American Community Survey
Design and Methodology
(January 2014)

Chapter 4:
Sample Design and Selection

[This page intentionally left blank]

Table of Contents

Figures

Tables

[This page intentionally left blank]

Chapter 4: Sample Design and Selection

4.1 Overview

The American Community Survey (ACS) and Puerto Rico Community Survey (PRCS) each consist of two separate samples: housing unit (HU) addresses and residents of group quarters (GQ) facilities. As described in Chapter 3, we derive the sampling frames from which we draw these samples from the Census Bureau's Master Address File (MAF). The MAF is the Census Bureau's official inventory of known living quarters and selected nonresidential units in the United States (U.S.) and Puerto Rico.

We select independent HU address samples for each of the 3,143 counties and county equivalents in the U.S., including the District of Columbia, as well as for each of the 78 municipalities in Puerto Rico. In 2004, we selected samples of HU addresses for every county and county equivalent for field data collection in 2005.[1] Each year from 2005–2010, we selected approximately 2.9 million HU addresses in the U.S. and 36,000 HU addresses in Puerto Rico. Beginning in 2011, we implemented the following changes to the ACS sample designs:

- We increased the HU sample in June 2011, bringing the size of the sample selected to 3.54 million addresses per year.
- We added several new HU sampling rates that better control the allocation of the sample and improve estimate reliability for small areas.
- We increased the follow-up sample to 100 percent in select geographic areas.

In addition, starting in 2013, we restricted the assignment of the GQ sample for college dorms to the non-summer months (January–April, September–December).

Full-implementation samples of GQ facilities and persons are selected independently within each state, including the District of Columbia and Puerto Rico. This began in 2006. In 2006 and 2007, the ACS and the PRCS included approximately 2.5 percent of the expected number of residents in GQ facilities. Beginning in 2008, we increased the sampling rates in 16 states with small GQ populations to meet publication thresholds. See Chapters 7 and 8 for details of the data collection methods.

This chapter presents details on the selection of the HU address and GQ samples. The final section describes the differences in sampling and data collection methodology for some hard to reach areas in Alaska (referred to as Remote Alaska). The section on Remote Alaska also details recently modified sampling and data collection procedures for these areas.

---

[1]In the remainder of this chapter, the term "county" refers to counties, county equivalents, and municipalities.

4.2 Housing Unit Sample Selection

There are two phases of HU address sampling for each county.[2] During first-phase sampling, we assign blocks to sampling strata, calculate sampling rates, and select the sample. During the second phase of sampling, we select a sample of nonresponding addresses for Computer Assisted Personal Interviewing (CAPI). This is the CAPI sample.

First-phase sampling produces the annual ACS initial sample of addresses and includes two processes—main and supplemental sampling. The main and supplemental samples in the first-phase sampling include two stages. The first stage sample selection systematically assigns new addresses to sub-frames and identifies the appropriate sub-frame associated with a specific year's sample. The second stage sample selection systematically selects the sample from the selected sub-frame.

Figure 4-1 provides a visual overview of the housing unit address sampling process.

---

[2]Throughout this chapter, "addresses" refers to valid ACS addresses that have met the filter criteria (Bates, Editing the MAF Extracts and Creating the Unit Frame Universe for the American Community Survey, 2013).

FIRST-PHASE SAMPLING

MAIN PROCESSING – SEPTEMBER/OCTOBER

Assign all blocks and addresses to sixteen sampling strata

Determine base rate and calculate stratum sampling rates

SUPPLEMENTAL PROCESSING - JANUARY

Match new addresses by block and assign to sampling strata

Determine base rate and calculate stratum sampling rates

FIRST-STAGE SAMPLE SELECTION
- Systematically assign new addresses to five existing sub-frames
- Identify sub-frame associated with current year

SECOND-STAGE SAMPLE SELECTION
- Systematically select sample from first-stage sample (sub-frame)

DATA COLLECTION

MAIL/INTERNET RESPONSES

CATI RESPONSES

NON-RESPONSES

SECOND-PHASE (CAPI) SAMPLE SELECTION - MONTHLY
- Select sample of unmailable addresses and non-responding addresses and send to CAPI
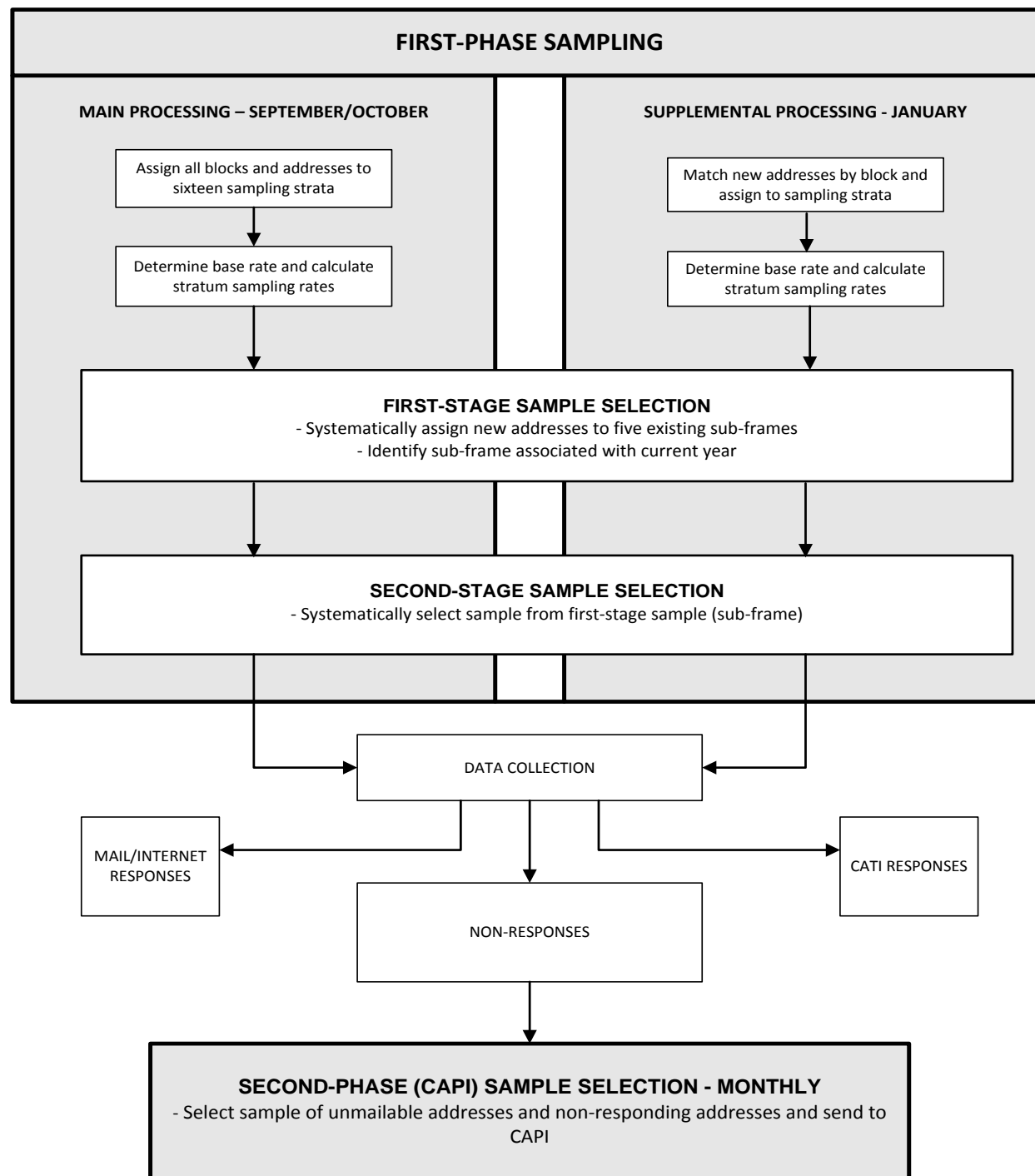
Figure 4-1: Assignment of Blocks (and their addresses) to Second-stage Sampling

4.3   First Phase Sample

The first phase of sampling is comprised of two separate stages. The first stage of first phase sampling maintains five distinct partitions, or sub-frames, of the addresses on the sampling frame within each county. Each county sub-frame is a representative sample of addresses in the county. We assign these sub-frames to specific years and rotate them annually. The sub-frames maintain their annual designation over time. First stage sampling systematically sorts and assigns addresses that are new to the frame to one of the five sub-frames.[3] First stage sampling also determines the sampling rates for each stratum for the current sample year.

The second stage of first phase sampling selects a sample of the addresses from the current year's sub-frame and allocates this sample to the twelve months of the year for data collection.

First-Phase, First-Stage Sample: Random Assignment of Addresses to a Specific Year

One of the ACS design requirements is that no HU address be in sample more than once in any five-year period. To accommodate this restriction, the addresses in the frame are assigned systematically to five sub-frames, each containing roughly 20 percent of the frame, and each being a representative sample. Addresses from only one of these sub-frames are eligible to be in the ACS sample each year and each sub-frame is used every fifth year. For example, 2014 will have the same addresses in its sub-frame as did 2009 with the addition of all new addresses that we assigned to that sub-frame during the 2010–2014 time period. As a result, we must perform both the main and supplemental sample selection in two stages. The first stage partitions the sampling frame into the five sub-frames and determines the sub-frame for the current year. The second stage, described in more detail below, selects addresses to be included in the ACS from the sub-frame eligible for the sample year.

Prior to the 2005 sample selection, there was a one-time allocation of all addresses then present on the ACS frame to the five sub-frames. In subsequent years, we must systematically allocate addresses new to the frame to these five sub-frames. We accomplish this by sorting the addresses in each county by stratum and geographic order including tract, block, street name, and house number. We then assign addresses sequentially to each of the five existing sub-frames. This procedure is similar to the use of a systematic sample with a sampling interval of five, in which the first address in the interval is assigned to year one, the second address in the interval to year two, and so on. Specifically, during main sampling, only the addresses new to the MAF since the previous year's supplemental MAF are eligible for first-stage sampling and go through the process of assignment to a sub-frame. Similarly, during supplemental sampling, only addresses new to the MAF since main sampling go through first-stage sampling.

---

[3]All existing addresses retain their previous assignment to one of the five sub-frames. The five sub-frames are maintained to meet the requirement that no address be in sample more than once in a five-year period.

The ACS and PRCS reflect two separate sampling operations carried out at different times of the year: (1) main sampling, which occurs in September and October of the year preceding the sample year, and (2) supplemental sampling, which occurs in January of the sample year. This allows an opportunity for new addresses to have a chance of selection into the sample. The ACS sampling frames for both main and supplemental sampling are derived from the most recently updated MAF, so the sampling frames for the main and supplemental sample selections differ for a given sample year. The MAF available at the time of main sampling, obtained in the July preceding the sample year, reflects address updates through March of that year. The MAF available at the time of the supplemental sample selection, obtained in January of the sample year, reflects address updates through September of the year preceding the sample year. During supplemental sampling, we assign addresses new to the frame systematically to the five sub-frames using the same process for new addresses as in the main sample.

First Phase, First-Stage Sample: Determining the Sampling Rates

Each year, we must determine the specific set of sampling rates for each of the thirteen non-fixed rate sampling strata defined in Table 4-1. Before we can do this, we must perform the following two steps. The first step is to calculate a base rate (BR) for the current year. Thirteen of the sixteen sampling rates are a function of a base rate. The three fixed rate strata are 15 percent, 10 percent, and 7 percent. Column 3 of Table 4-1 shows the relationship between the base rate and the sixteen sampling rates. Beginning in 2009, the number of new addresses differed from what was expected by enough to warrant the calculation of a separate set of sampling rates for supplemental sample selection. This led to separate supplemental sampling rates beginning with the 2010 sample selection.

The distribution of addresses by sampling stratum, coupled with the target sample size of 3.54 million, allows us to set up and solve a simple algebraic equation for the BR.

The second step is the calculation of the sampling rates using the value of BR and the equations in Table 4-1. Beginning in June, 2011 we increased the sample size to a monthly level corresponding to an annual 3.54 million sample (approximately 295,000 per month). Between January of 2005 and May of 2011, the monthly sample corresponded to an annual sample of approximately 2.9 million (roughly 242,000 per month).

First-Phase, First-Stage Sample: First-Phase Sampling Rates

Columns 2 and 3 of Table 4-1 provide the sampling rates for the 2013 ACS for the U.S. and Puerto Rico, respectively (Sommers, 2012b).

Table 4-1: 2013 ACS/PRCS Main Sampling Rates

Stratum	Sampling Rates[1]	
	United States	Puerto Rico
Blocks in smallest sampling entities (0<SEMOS≤200)	15.0	(NA)
Blocks in small sampling entities (200<SEMOS≤400)	10.0	(NA)
Blocks in medium sampling entities (400<SEMOS≤800)	7.0	7.0
Blocks in large sampling entities (800<SEMOS≤1,200)	4.4	(NA)
Blocks in large sampling entities (SEMOS>1,200) and smallest tracts (0<TMOS≤400) with predicted levels of completed interviews prior to CAPI sampling≤60 percent	5.5	4.9
Blocks in large sampling entities (SEMOS>1,200) and smallest tracts (0<TMOS≤400) with predicted levels of completed interviews prior to CAPI sampling>60 percent	5.1	
Blocks in large sampling entities (SEMOS>1,200) and small tracts (400<TMOS≤1,000) with predicted levels of completed interviews prior to CAPI sampling≤60 percent	4.4	3.9
Blocks in large sampling entities (SEMOS>1,200) and small tracts (400<TMOS≤1,000) with predicted levels of completed interviews prior to CAPI sampling>60 percent	4.0	
Blocks in large sampling entities (SEMOS>1,200) and medium tracts (1,000<TMOS≤2,000) with predicted levels of completed interviews prior to CAPI sampling≤60 percent	2.7	2.4
Blocks in large sampling entities (SEMOS>1,200) and medium tracts (1,000<TMOS≤2,000) with predicted levels of completed interviews prior to CAPI sampling>60 percent	2.5	
Blocks in large sampling entities (SEMOS>1,200) and large tracts (2,000<TMOS≤4,000) with predicted levels of completed interviews prior to CAPI sampling≤60 percent	1.6	1.4
Blocks in large sampling entities (SEMOS>1,200) and large tracts (2,000<TMOS≤4,000) with predicted levels of completed interviews prior to CAPI sampling>60 percent	1.4	
Blocks in large sampling entities (SEMOS>1,200) and larger tracts (4,000<TMOS≤6,000) with predicted levels of completed interviews prior to CAPI sampling≤60 percent	0.9	0.8
Blocks in large sampling entities (SEMOS>1,200) and larger tracts (4,000<TMOS≤6,000) with predicted levels of completed interviews prior to CAPI sampling>60 percent	0.9	
Blocks in large sampling entities (SEMOS>1,200) and largest tracts (6,000>TMOS) with predicted levels of completed interviews prior to CAPI sampling≤60 percent	0.5	(NA)
Blocks in large sampling entities (SEMOS>1,200) and largest tracts (6,000>TMOS) with predicted levels of completed interviews prior to CAPI sampling>60 percent	0.5	

Note: The rates in the table have been rounded to one decimal place.

NA Not applicable.

[1] In percent.

Since the design of the ACS calls for a target annual address sample of approximately 3.54 million in the U.S. and 36,000 in Puerto Rico, we reduce the sampling rates for all but the smallest sampling entity strata (SEMOS≤800) each year as the number of addresses in the U.S. and Puerto Rico increases. However, as shown in Table 4-1, among the strata where the rates are decreasing, the relationship of the sampling rates will remain proportionally constant. The sampling rates for the smallest sampling entities will remain at 15 percent, 10 percent, and 7 percent.

The sampling rates that we use to select the sample include strata for blocks in certain census tracts in the U.S. These tracts are projected to have the highest rates of completed questionnaires by mail and by the telephone follow-up operation, called Computer Assisted Telephone Interviewing (CATI). This adjustment is to compensate for the increase in costs due to increasing the CAPI sampling rates in tracts predicted to have the lowest rate of completed interviews by mail and CATI. Note that the initial identification of these tracts, performed in 2004 was used in the 2005 sample selection and was revised in 2007 based on more recent data and has been used since the 2008 sample selection.

Specifically, we multiply the sampling rates by 0.92 (reduced by 8 percent) for blocks in the U.S. in the six strata in which the SEMOS was greater than 1,200. We make this adjustment for blocks in tracts that we predict will have a level of completed mail and CATI interviews of at least 60 percent, and at least 75 percent mailable addresses.

Because of this adjustment, there are sixteen sampling rates used in the U.S., and ten in Puerto Rico, as shown in columns 2 and 3 of Table 4-1. See the research report (Asiala, 2005) for a full description of the relationship between this reduction and the CAPI sampling rates. This reduction does not occur in Puerto Rico, therefore there are ten sampling strata eligible to be used in Puerto Rico. Only six strata in Puerto Rico contain valid addresses on the 2013 main sampling frame, so for 2013, we only used the six sampling rates shown in Table 4-1.

First-Phase, Second-Stage Sampling: Selection of Addresses

As noted earlier, the second stage of first phase sampling selects a sample of the addresses from the current year's sub-frame. We partition this sub-frame by county and select the addresses from the sub-frame in each county. Second stage sampling allocates this sample to the twelve months of the year for data collection. This process results in the creation of the initial annual ACS sample.

We sort the addresses in each county by stratum and the first-stage order of selection. After sorting, we select systematic samples of addresses using a sampling rate approximately equal to the final sampling rate divided by 20 percent.[4]

First-Phase, Second-Stage Sampling: Assigning Addresses to the Second-Stage Sampling Strata

Each year, the main sampling operation assigns each block to one of the sixteen sampling strata, and consequently, assigns each block one of sixteen sampling rates.[5] The ACS produces estimates for geographic areas having a wide range of population sizes. To ensure that the estimates for these areas have the desired level of reliability, we must sample areas with smaller populations at higher rates relative to those areas with larger populations. We base the stratum assignment for a block on information about the set of geographic entities—referred to as sampling entities—which contain the block, or on information about the size of the census tract that the block is located in, as discussed below. Sampling entities are:

- Counties,

- Places with active and functioning governments,[6]

- School districts,

- American Indian Areas/Alaska Native Areas/Hawaiian Home Lands (AIANHH),

- American Indian Tribal Subdivisions with active and functioning governments,

- Minor civil divisions (MCDs) with active and functioning governments in 12 states,[7] or

- Census Designated Places (CDPs) in Hawaii only.

We base the sampling stratum for most blocks on the measure of size (MOS) for the smallest sampling entity to which any part of the block belongs. To calculate the MOS for a sampling entity, we derive block-level counts of addresses from the main MAF. This count is converted to an estimated number of occupied HUs by multiplying it by the proportion of occupied HUs in the block in the 2010 Census. For American Indian and Alaska Native Statistical Areas

---

[4]The second-stage rate is approximately equal to the sampling rate divided by 20 percent since the first-stage sampling rate is approximately 20 percent, and the first-stage rate times the second-stage rate equals the overall sampling rate. An adjustment is made to account for uneven distributions of addresses in the county level sub-frames.

[5] From 2005 – 2010 five sampling strata were used.

[6] Functioning governments have elected officials who can provide services and raise revenue.

[7] The 12 states are considered "strong" MCD states and are: Connecticut, Maine, Massachusetts, Michigan, Minnesota, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont, and Wisconsin.

(AIANSA[8]) and Tribal Subdivisions, we multiply the estimated number of occupied HUs by the proportion of its population that responded as American Indian or Alaska Native (either alone or in combination) in the 2010 Census.[9] For each sampling entity, we sum the estimate across all blocks in the entity to create the MOS for the entity. In AIANSAs, if the sum of these estimates across all blocks is non-zero, then this sum becomes the MOS for the AIANSA. If it is zero (due to a zero census count of American Indians or Alaska Natives), the occupied HU estimate for the AIANSA is the MOS for the AIANSA. For detail, see the computer specifications for calculating the MOS for the ACS (Sommers, 2012a). We assign each block the smallest MOS of all the sampling entities in which the block is contained and we refer to it as the Smallest Entity Measure of Size, or SEMOS.

If the SEMOS is greater than 1,200, we base the stratum assignment for the block on the MOS for the census tract that contains it. The sum of the estimated number of occupied HUs across all of its blocks.is the MOS for each tract (TMOS). Using SEMOS and TMOS, we can assign blocks to the sixteen strata defined in columns 1 and 2 in Table 4-2 below.

---

[8] AIANSA is a general term used to describe American Indian and Alaska Native Village statistical areas. For detailed technical information on the Census Bureau's American Indian and Alaska Native Areas Geographic Program for Census 2000, see the publication in the *Federal Register* (U.S. Census Bureau, 2000).

[9] 2010 Census information was used for the first time to define the measures of size in the 2012 sample selection.

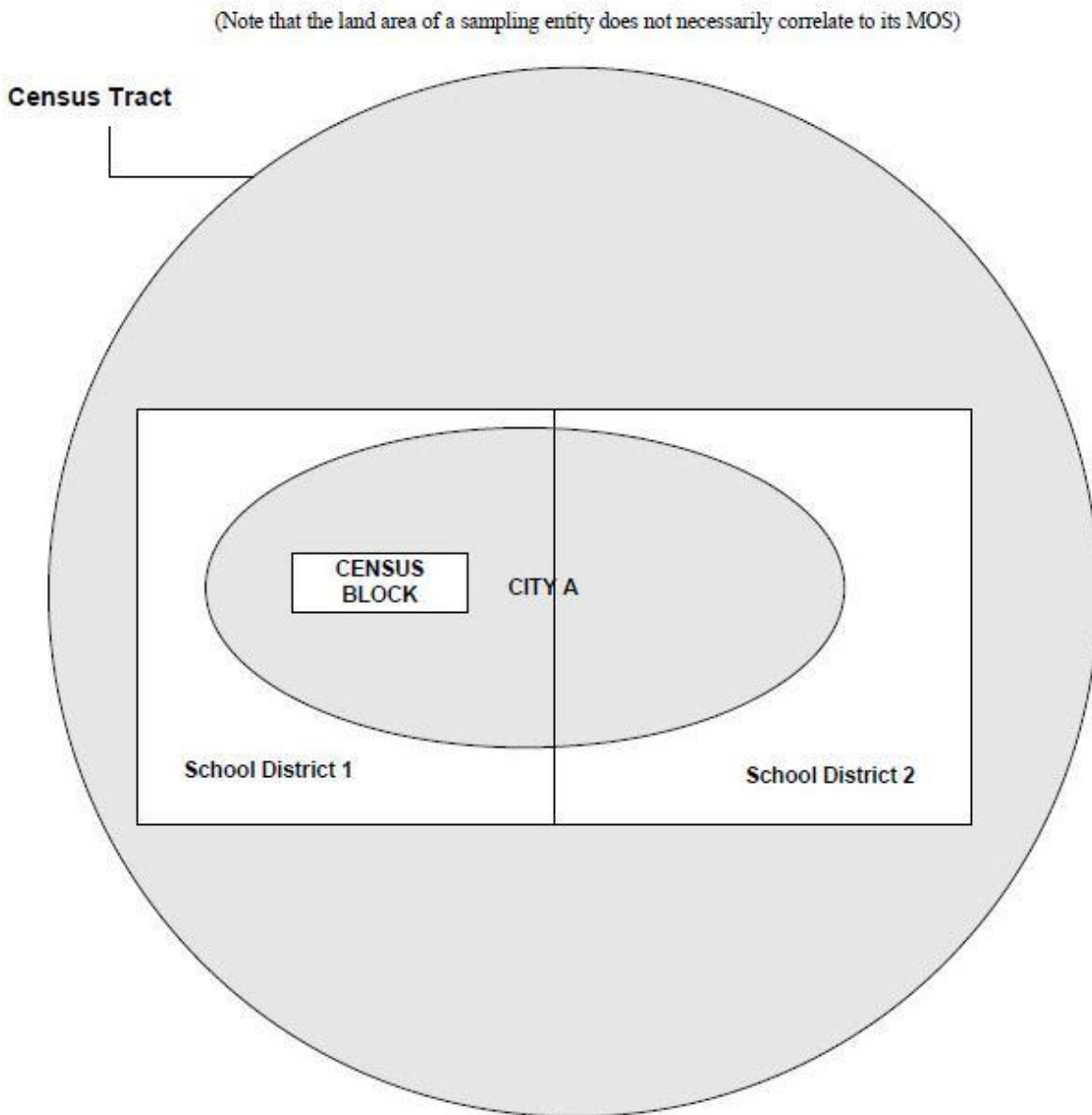**Table 4-2: Sampling Strata Thresholds and Relationship between the Base Rate and the Sampling Rates**

Stratum	Smallest Entity Measure of Size (SEMOS) and Tract Measure of Size (TMOS)	Sampling Rates
Blocks in smallest sampling entities	$0 < \text{SEMOS} \leq 200$	15% (fixed)
Blocks in small sampling entities	$200 < \text{SEMOS} \leq 400$	10% (fixed)
Blocks in medium sampling entities	$400 < \text{SEMOS} \leq 800$	7% (fixed)
Blocks in large sampling entities	$800 < \text{SEMOS} \leq 1,200$	$2.8 \times \text{BR}$
Blocks in large sampling entities (SEMOS > 1,200) and smallest tracts with predicted levels of completed interviews prior to CAPI sampling ≤ 60 percent	$0 < \text{TMOS} \leq 400$	$3.5 \times \text{BR}$
Blocks in large sampling entities (SEMOS > 1,200) and smallest tracts with predicted levels of completed interviews prior to CAPI sampling > 60 percent		$0.92 \times 3.5 \times \text{BR}$
Blocks in large sampling entities (SEMOS > 1,200) and small tracts with predicted levels of completed interviews prior to CAPI sampling ≤ 60 percent	$400 < \text{TMOS} \leq 1,000$	$2.8 \times \text{BR}$
Blocks in large sampling entities (SEMOS > 1,200) and small tracts with predicted levels of completed interviews prior to CAPI sampling > 60 percent		$0.92 \times 2.8 \times \text{BR}$
Blocks in large sampling entities (SEMOS > 1,200) and medium tracts with predicted levels of completed interviews prior to CAPI sampling ≤ 60 percent	$1,000 < \text{TMOS} \leq 2,000$	$1.7 \times \text{BR}$
Blocks in large sampling entities (SEMOS > 1,200) and medium tracts with predicted levels of completed interviews prior to CAPI sampling > 60 percent		$0.92 \times 1.7 \times \text{BR}$
Blocks in large sampling entities (SEMOS > 1,200) and large tracts with predicted levels of completed interviews prior to CAPI sampling ≤ 60 percent	$2,000 < \text{TMOS} \leq 4,000$	BR
Blocks in large sampling entities (SEMOS > 1,200) and large tracts with predicted levels of completed interviews prior to CAPI sampling > 60 percent		$0.92 \times \text{BR}$
Blocks in large sampling entities (SEMOS > 1,200) and larger tracts with predicted levels of completed interviews prior to CAPI sampling ≤ 60 percent	$4,000 < \text{TMOS} \leq 6,000$	$0.6 \times \text{BR}$
Blocks in large sampling entities (SEMOS > 1,200) and larger tracts with predicted levels of completed interviews prior to CAPI sampling > 60 percent		$0.92 \times 0.6 \times \text{BR}$
Blocks in large sampling entities (SEMOS > 1,200) and largest tracts with predicted levels of completed interviews prior to CAPI sampling ≤ 60 percent	$6,000 < \text{TMOS}$	$0.35 \times \text{BR}$
Blocks in large sampling entities (SEMOS > 1,200) and largest tracts with predicted levels of completed interviews prior to CAPI sampling > 60 percent		$0.92 \times 0.35 \times \text{BR}$

Figure 4-2 shows a Census Block that is in City A and contained in school district 1. Therefore, it is contained wholly in three sampling entities:

- County (not shown)
- Place with active and functioning government—City A
- School district

Example 1: Suppose the MOS for City A is 600 and the MOS for School District 1 is 1,100. Then the SEMOS for the Census Block is 600 and it is placed in the $400 < \text{SEMOS} \leq 800$ stratum.

Example 2: Suppose the MOS for City A is 1,300 and the MOS for School District 1 is 1,400.Then the SEMOS for the block is 1,300. Since the SEMOS for the block is greater than 1,200 the block will be assigned to one of the twelve strata with $\text{SEMOS} > 1,200$ depending on the size of the census tract (TMOS - not shown in the diagram) and the predicted level of completed interviews prior to CAPI sampling in the tract. In this example, suppose the TMOS is 1,800, and the predicted level of completed interviews prior to CAPI sampling is ≤ 60 percent, then the Census Block will be placed in the $1,000 < \text{TMOS} \leq 2,000$ stratum with a predicted level of completed interviews prior to CAPI sampling ≤ 60 percent.

(Note that the land area of a sampling entity does not necessarily correlate to its MOS)

Census Tract

CENSUS BLOCK

CITY A

School District 1

School District 2

Figure 4-2: Assignment of Blocks (and their addresses) to Second-stage Sampling

First-Phase, Second-Stage Sampling: Sample Month Assignment for Address Samples

We must assign each sample address for a particular year to a specific data collection month. The set of all addresses assigned to a specific month is the month's sample or panel. We sort addresses selected during main sampling by stratum and geography and assign them systematically to the 12 months of the year. However, we assign addresses that one of several Census Bureau household surveys have also selected, to an ACS data collection month based on the interview month(s) for these other household surveys.[10] The goal of the assignments is to reduce the respondent burden of completing interviews for both the ACS and another survey during the same month.

We sort the supplemental sample by stratum and geography and systematically assign this sample to the months of July through December. Since this sample is only approximately one percent of the total ACS sample, very few addresses are also in one of the other household surveys in the specified months. Therefore, we chose not to implement the procedure described above to move the ACS data collection month for cases in common with the current surveys during supplemental first-phase sampling.

4.4   Second-Phase Sampling for CAPI follow-up

The ACS uses four modes of data collection—Internet, mail, telephone, and personal visit. (See Chapter 7 for more information on data collection.) Mailable sample addresses are eligible to complete the survey during the entire three-month time period. We send all mailable addresses with available telephone numbers for which we receive no Internet or mail response during the first data collection month to CATI for follow-up. We conduct CATI follow-up for these cases during the second month. Cases without a completed Internet or mail questionnaire or a completed CATI interview are eligible for CAPI in the third month, as are the unmailable addresses. An address is unmailable if the address is incomplete or directs mail to only a post office box. Table 4-3 summarizes the eligibility of addresses for CAPI sampling.

---

[10]These surveys include the Survey of Income and Program Participation, the National Crime Victimization Survey, the Consumer Expenditures Quarterly and Diary Surveys, the Current Population Survey, and the State Child Health Insurance Program Surveys.

Table 4-3: Addresses Eligible for CAPI Sampling

Mailable Address	Responds to Mailing (Internet or questionnaire)	Responds to CATI	Eligible for CAPI
No	(NA)	(NA)	Yes
Yes	No	No	Yes
Yes	No	Yes	No (completed)
Yes	Yes	(NA)	No (completed)

NA not applicable

The CAPI sample selects a systematic sample of these addresses for CAPI data collection each month using the rates shown in Table 4-4. The selection is made after sorting within county by CAPI sampling rate, mailable versus unmailable, and geographic order within the address frame (Keathley, 2010).

The variance of estimates for HUs and people living in them in a given area is a function of the number of interviews completed within that area. However, due to the subsampling, CAPI cases generally have larger weights than cases completed by Internet, mail or CATI. The variance of the estimates for an area will tend to increase as the proportion of Internet, mail, and CATI responses decreases. Large differences in these proportions across areas of similar size may result in substantial differences in the reliability of their estimates. To minimize this possibility, tracts in the U.S. that are predicted to have low levels of interviews completed by Internet, mail and CATI have their CAPI sampling rates adjusted upward from the default 1-in-3 rate for mailable addresses. This tends to reduce variances for the affected areas both by potentially increasing their total numbers of completed interviews and by decreasing the differences in weights between their CAPI interviews and mail/Internet/CATI interviews.

No information was available to reliably predict the levels of completed interviews prior to second-phase sampling for CAPI follow-up in Puerto Rico prior to 2005, so we initially used the sampling rates of 1-in-3 for mailable and 2-in-3 for unmailable addresses. On the basis of early response results observed during the first months of the PRCS, we changed the CAPI sampling rate for mailable addresses in all Puerto Rico tracts to 1-in-2 beginning in June 2005.

We made several enhancements to the CAPI sampling beginning with the 2011 sample, to increase the reliability of the ACS estimates for populations in certain well-defined geographic areas. Beginning in January of 2011, we send all Remote Alaska sample addresses to CAPI where previously they had been sampled at the rate assigned to unmailable cases (2-in-3). In addition, we send all unmailable addresses and all addresses that did not respond via Internet, mail, or CATI to CAPI in the following areas:

- Hawaiian Homelands
- Alaska Native Village Statistical Areas
- All American Indian areas with at least ten percent of the population responding to the 2010 Census as American Indian or Alaska Native (alone or in combination).

Table 4-4 summarizes the CAPI sampling rates that are used for addresses of each particular type.

Table 4-4: CAPI Sampling Rates

Address and Tract Characteristics	CAPI Sampling rate (percent)
United States	
Addresses in Remote Alaska, mailable and unmailable addresses in American Indian areas with 10 percent or more American Indian population (alone or in combination) in the 2010 Census, mailable and unmailable addresses in Hawaiian Homelands, mailable and unmailable addresses in Alaska Native Village Statistical Areas	100.0
Other unmailable addresses	66.7
Mailable addresses in tracts with predicted levels of completed interviews prior to CAPI subsampling between 0 percent and 35 percent	50.0
Mailable addresses in tracts with predicted levels of completed interviews prior to CAPI subsampling greater than 35 percent and less than 51 percent	40.0
Mailable addresses in other tracts	33.3
Puerto Rico	
Unmailable addresses	66.7
Mailable addresses	50.0

4.5 Group Quarters Sample Selection

GQ facilities include such places as college residence halls, residential treatment centers, skilled nursing facilities, group homes, military barracks, correctional facilities, workers' dormitories, and facilities for people experiencing homelessness. We classify each GQ facility according to its GQ type. (For more information on GQ facilities, see Chapter 8.) As noted previously, the 2005 ACS did not include GQ facilities, but we have included GQs since 2006. We select the GQ sample for a given year during a single operation carried out in September and October of the previous year. The most recently available updated MAF as well as lists from other sources and operations define the sampling frame of GQ facilities and their locations. The ultimate sampling units for the GQ sample are the GQ residents, not the facilities. The GQ samples are independent state-level samples.

The ACS sampling and data collection operations exclude certain GQ types , including domestic violence shelters, soup kitchens, regularly scheduled mobile food vans, targeted non-sheltered outdoor locations, commercial maritime vessels, natural disaster shelters, and dangerous encampments. There are several reasons for their exclusion and they vary by GQ type. Concerns

about privacy and the operational feasibility of repeated interviewing for a continuing survey, rather than once a decade for a census, led to the decision to exclude these GQ types. However, we control ACS estimates of the total population to be consistent with the Population Estimates Program estimate of the GQ resident population from all GQs, even those excluded from the ACS.

We classify all GQ facilities into one of two groups: (1) small GQ facilities (having 15 or fewer people according to 2010 Census or updated information); (2) large GQ facilities (with an expected population of more than 15 people). There are approximately 94,000 small GQ facilities and 68,000 large GQ facilities on the 2013 GQ sampling frame. We create two sampling strata to sample the GQ facilities. The first stratum includes both small GQ facilities and those with no available population count. The second stratum includes large facilities. In the remainder of this chapter, these strata will be referred to as the small GQ stratum and the large GQ stratum. We compute a GQ measure of size (GQMOS) for use in sampling the large GQ facilities. The GQMOS for each large GQ is the expected population count divided by 10.

The sampling procedures differ for these two strata. We sample GQs in the small GQ stratum like addresses in the HU sample, and collect data for all people in the selected GQ facilities. Like HU addresses, small GQ facilities are eligible to be in the sample only once in a five-year period. People are the ultimate sampling unit for GQs in the large stratum, where groups of 10 people ("hits") are selected for interview from GQ facilities in the large GQ stratum, and the number of these groups selected for a large GQ facility is a function of its GQMOS. Unlike HU addresses and small GQs, large GQ facilities are eligible for sampling each year. For more detail, see the computer specifications for the GQ sampling (Cyffka, 2012).

4.6   Small Group Quarters Stratum Sample

For the small GQ stratum, a two-phase, two-stage sampling procedure is used. In the first phase, we select a GQ facility sample using a method similar to that used for the first-phase HU address sample. Just as we saw in the HU address sampling, the first phase has two stages. Stage 1 systematically assigns small GQ facilities to a sub-frame associated with a specific year. During the second stage, we select a systematic sample of the small GQ facilities. In the second phase of sampling, we interview all people in the facility as long as there are 15 or fewer at the time of interview. Otherwise, we select and interview a sub-sample of 10 people.

First Phase of Small GQ Sampling—Stage One: Random Assignment of GQ Facilities to Sub-frames

The sampling procedure for 2006 assigned all of the GQ facilities in the small stratum to one of five 20 percent sub-frames. We sort the GQ facilities within each state by small versus closed on Census Day, new versus previously existing, GQ type (such as skilled nursing facility, military barracks, or dormitory), and geographical order (county, tract, block, street name, and GQ identifier) in the small GQ frame. In each year subsequent to 2006, we assigned new GQ

facilities systematically to the five sub-frames. The sub-frame for the 2013 GQ sample selection contains the facilities previously designated to the sub-frame for calendar year 2013 and 20 percent of new small GQ facilities added since the 2012 sampling activates. The small GQ facilities in the 2013 sub-frame will not be eligible for sampling again until 2018, since the once-in-five-years period restriction also applies to small GQ facilities.

First Phase of Small GQ Sampling—Stage Two: Selection of Facilities

The second-stage sample is a systematic sample of the GQ facilities from the assigned sub-frame within each state. The GQs are sorted by new versus previously existing addresses and the order in which they were selected during stage one sampling. Regardless of their actual size, all of these small GQ facilities have the same probability of selection. The second-stage sampling rate combined with the 1-in-5 first-stage sampling rate yields an overall first-phase-sampling rate equal to the sampling rate for each state. As an example, if the sampling rate for the state is 2.5

Table 4-5: 2012 Group Quarters State-level Sampling Rates

State	Sampling Rate (percent)	State	Sampling Rate (percent)
Alabama	2.17	Montana	3.96
Alaska	4.19	Nebraska	2.46
Arizona	2.05	Nevada	3.63
Arkansas	2.21	New Hampshire	2.90
California	2.49	New Jersey	2.72
Colorado	2.33	New Mexico	2.77
Connecticut	2.37	New York	2.29
Delaware	5.00	North Carolina	2.34
District of Columbia	2.77	North Dakota	4.49
Florida	2.34	Ohio	2.39
Georgia	2.39	Oklahoma	2.39
Hawaii	3.00	Oregon	2.50
Idaho	4.13	Pennsylvania	2.53
Illinois	2.21	Rhode Island	2.63
Indiana	2.35	South Carolina	2.26
Iowa	2.40	South Dakota	3.51
Kansas	2.39	Tennessee	2.30
Kentucky	2.38	Texas	2.12
Louisiana	2.60	Utah	3.00
Maine	3.09	Vermont	4.39
Maryland	2.39	Virginia	2.20
Massachusetts	2.22	Washington	2.45
Michigan	2.79	West Virginia	2.31
Minnesota	2.47	Wisconsin	2.47
Mississippi	2.32	Wyoming	6.97
Missouri	2.25	Puerto Rico	2.50

percent, then the second-stage sampling rate would be 1-in-8 so that overall the GQ sampling would be (1-in-5) × (1-in-8) = 1-in-40 = 2.5 percent. Table 4-5 shows the 2012 state level sampling rates.

Second Phase of Small GQ Sampling: Selection of Persons within Selected Facilities

Every person in the GQ facilities selected in this sample is eligible to be interviewed. If the number of people in the GQ facility exceeds 15, interviewers perform a field sub-sampling operation to reduce the total number of sampled people to 10, similar to the groups of ten selected in the large GQ stratum.

4.7   Large Group Quarters Stratum Sample

Unlike the HU address and small GQ samples, we do not divide the large GQ facilities into five sub-frames. The ultimate sampling units for large GQ facilities are people, not the facility itself, and we conduct interviews in groups of ten. We use a two-phase sampling procedure. The first phase indirectly selects the GQ facilities by selecting groups of ten within the facilities. The second phase selects the people for each facility's group(s) of ten. The number of groups of ten eligible to be sampled from a large GQ facility is equal to its GQMOS. For example, if a facility had 550 people in the 2010 Census, its GQMOS is 55 and there are 55 groups of ten that are eligible for selection in the sample.

First Phase of Large GQ Sampling: Selection of Groups of Ten (and Associated Facilities)

We sort all of the large GQ facilities in a state by GQ type and geographical order in the large GQ frame, and select a systematic sample of groups of ten. For this reason, in states with a 2.5 percent sampling rate, a GQ facility with fewer than 40 groups (or roughly 400 individuals) may or may not have one of its groups selected for the sample. GQ facilities in a state with a 2.5 percent sampling rate and between 40 and 80 groups will have at least one group selected with certainty. If the GQ facility has between 80 and 120 groups, it will have at least two groups selected and so forth.

Second Phase of Large GQ Sampling: Selection of Persons within Facilities

The second phase of sampling takes place within each large GQ facility that has at least one group selected in the first stage. When a field representative visits a GQ facility to conduct interviews, an automated listing instrument randomly selects the 10 people to be included, one from each group of ten being interviewed. The instrument is pre-loaded with the number of expected person interviews (ten times the number of groups selected) and a random starting number. The field representative then enters the actual number of people in the facility, as well as a roster of their names. To achieve a group size of 10, the instrument computes the appropriate sampling interval based on the observed population at the time of interviewing and then selects the actual people for interviewing using a pre-loaded random start and a systematic algorithm. If

the large GQ has an observed population of 15 or fewer people, the instrument selects a group size of 10; if the observed population is less than 10, the instrument selects everyone in the GQ.

For most GQ types, if multiple groups are selected within a GQ facility, their groups of ten are assigned to different sample months for interviewing. Very large GQ facilities with more than 12 groups selected have multiple groups assigned to some sample months. In these cases, we try to avoid selecting the same person more than once in a sample month. However, there is no attempt made to avoid selection of someone more than once across sample months within a year. Thus, we could interview someone in a very large GQ facility in consecutive months. All GQ facilities in this stratum are eligible for selection every year, regardless of their sample status in previous years.

Sample Month Assignment for Small and Large Group Quarter Samples

We assign the selected small GQ facilities and groups of ten for large GQ facilities to months using a procedure similar to the one used for sampled HU addresses. We combine and sort all GQ samples from a state by small versus large stratum and first-phase order of selection. Consecutive samples are assigned to the 12 months in a pre-determined order, starting with a randomly determined month.

Due to operational and budgeting constraints, we assign the same month to all sample groups of ten within certain types of correctional GQs or military barracks. For example, we assign all samples in federal prisons to September, and data collection may take up to 4.5 months, an exception to the six weeks allowed for all other GQ types. For the samples in non-federal correctional facilities—state prisons, local jails, halfway houses, military disciplinary barracks, and other correctional institutions—or military barracks, individual GQ facilities are randomly assigned to months throughout the year.

Beginning with the 2013 GQ sample, we no longer assign college dorms to the months of May-August. This is in response to the relatively low interview rates at these GQs during the summer months. In 2013, we only assign college dorm GQs to January-April and September-December.

4.8   Remote Alaska Sample

Remote Alaska is a set of rural areas in Alaska that are difficult to access and for which all HU addresses are treated as unmailable. There are approximately 30,000 HU addresses and 500 GQs in Remote Alaska. Due to the difficulties in field operations during specific months of the year and the extremely seasonal population in these areas, data collection operations in Remote Alaska differ from the rest of the country. In both the main and supplemental HU address samples, the month assigned for each Remote Alaska HU address is based on the county, place, AIANSA, or block group (in that order) in which it is contained. We assign all designated addresses located in each of these geographical entities to either January or September in such a way as to balance workloads between the months and to keep groups of cases together geographically. We sort the addresses for each month by county and geographical order in the

address frame, and beginning in 2011, all sample addresses are sent directly to CAPI (bypassing mail, Internet, and CATI for the HU sample) in the appropriate month. We assign the GQ sample in Remote Alaska to January or September using the same procedure and allow up to four months to complete the HU and GQ data collection for each of the two data collection periods.

4.9   References

Asiala, M. (2005). American Comunity Survey Research Report: Differential Sub-Sampling in the Computer Assisted Personal Interview Sample Selection in Areas of Low Cooperation Rates. DSSD 2005 American Community Survey Documentation Memorandum Series #ACS05-DOC-2. Washington, DC: U.S. Census Bureau.

Bates, L. (2013). Editing the MAF Extracts and Creating the Unit Frame Universe for the American Community Survey. DSSD 2013 American Community Survey Universe Creation Memorandum Series #ACS13-UC-1. Washington, DC: U.S. Census Bureau.

Cyffka, K. (2012). Specifications for Selecting the American Community Survey Group Quarters Sample. DSSD 2013 American Community Survey Sampling Memoradum Series #ACS13-S-6. Washington, DC: U.S. Census Bureau.

Keathley, D. (2010). American Community Survey: Specifications for Selecting the Computer Assisted Personal Interview Samples. DSSD 2010 American Community Survey Sampling Memorandum Series #ACS10-S-45. Washington, DC: U.S. Census Bureau.

Sommers, D. (2012a). Creating the Governmental Unit Measure of Size (GUMOS) Datasets for the American Community Survey and the Puerto Rico Community Survey. DSSD 2013 American Community Survey Sampling Memorandum Series #ACS13-S-1. Washington, DC: U.S. Census Bureau.

Sommers, D. (2012b). Specifications for Selecting the Main and Supplemental Housing Unit Address Samples for the American Community Survey. DSSD 2013 American Community Survey Sampling Memoradum Series #ACS13-S-3. Washington, DC: U.S. Census Bureau.

U.S. Census Bureau. (2000). American Indian and Alaska Native Areas Geographic Program for Census 2000; Notice. Federal Register , 65 (121), 39062-39069.